

DISK CONTROL SYSTEM, DISK CONTROL APPARATUS, DISK SYSTEM AND CONTROL METHOD THEREOF

Publication number: JP2003323261

Publication date: 2003-11-14

Inventor: KANAI HIROKI; KANEKO SEIJI

Applicant: HITACHI LTD

Classification:

- international: G06F12/08; G06F3/00; G06F3/06; G06F13/00;
G06F13/12; G06F12/08; G06F3/00; G06F3/06;
G06F13/00; G06F13/12; (IPC1-7): G06F3/06;
G06F12/08; G06F13/12

- european:

Application number: JP20020126885 20020426

Priority number(s): JP20020126885 20020426

Also published as:



US6961788 (B2)

US2003204649 (A1)

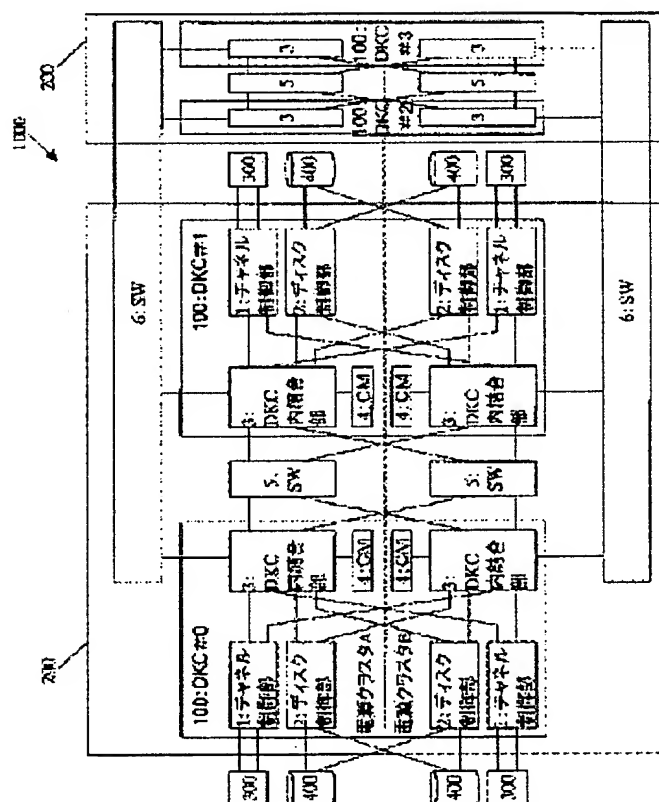
Report a data error here

Abstract of JP2003323261

PROBLEM TO BE SOLVED: To provide a control apparatus for effectively configuring with an identical architecture from a small scale configuration to super large scale configuration.

SOLUTION: A disk control unit comprises one or a plurality of channel control units having an interface with a host computer, one or a plurality of disk control units having an interface with a disk apparatus, and an internal connection unit for connecting a cache memory unit for temporarily storing data to be read/written from/to the disk apparatus, a channel control unit and a disk control unit. A first connection portion for connecting internal connection portions of disk control units for reading/writing data inside each of the disk control apparatuses, and a second connection portion for connecting internal connection portions of the disk control units for transferring data straddling the plurality of disk control apparatuses.

COPYRIGHT: (C)2004,JPO



Data supplied from the esp@cenet database - Worldwide

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2003-323261

(P2003-323261A)

(43) 公開日 平成15年11月14日 (2003. 11. 14)

(51) Int.Cl. ⁷	識別記号	P I	テマコード* (参考)
G 0 6 F 3/06	3 0 1 3 0 2	G 0 6 F 3/06	3 0 1 B 5 B 0 0 5 3 0 2 A 5 B 0 1 4 3 0 2 B 5 B 0 6 5
12/06	5 0 1 5 5 7	12/06	5 0 1 E 5 5 7
審査請求 未請求 請求項の数16 O L (全 10 頁) 最終頁に続く			

(21) 出願番号 特願2002-126885 (P2002-126885)

(22) 出願日 平成14年4月26日 (2002. 4. 26)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 金井 宏樹

神奈川県小田原市中里322番地2号 株式会社日立製作所RAIDシステム事業部内

(72) 発明者 金子 誠司

神奈川県小田原市中里322番地2号 株式会社日立製作所RAIDシステム事業部内

(74) 代理人 100071283

弁理士 一色 健輔 (外4名)

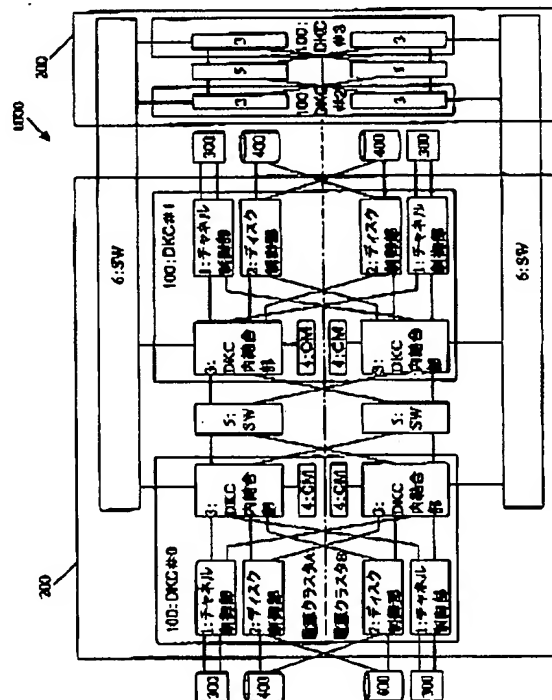
最終頁に続く

(54) 【発明の名称】 ディスク制御システム、ディスク制御装置、ディスクシステム、及びその制御方法

(57) 【要約】 (修正有)

【課題】 小規模な構成から超大規模な構成まで同一のアーキテクチャで効率よく構成することを可能とする制御装置を提供する。

【解決手段】 ディスク制御ユニットは、ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、ディスク装置とのインターフェースを有する一または複数のディスク制御部と、ディスク装置にリード/ライトされるデータを一時的に格納するキャッシュメモリ部とチャネル制御部とディスク制御部とを相互に接続する内部結合部とを備え、各ディスク制御装置の内部において、データをリード/ライトすべく、各ディスク制御ユニットの内部結合部を相互に結合する第一の結合部と、複数のディスク制御装置に跨り、データを転送すべく、各ディスク制御ユニットの内部結合部を相互に結合する第二の結合部とを備えたものとする。



(2)

特開2003-323261

1

2

【特許請求の範囲】

【請求項1】 複数のディスク制御ユニットを有するディスク制御装置を複数備えたディスク制御システムにおいて、

前記ディスク制御ユニットは、

ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、

ディスク装置とのインターフェースを有する一または複数のディスク制御部と、

前記ディスク装置にリード/ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャネル制御部と前記ディスク制御部とを相互に接続する内部結合部と、

を備えており、

前記各ディスク制御装置の内部において、データをリード/ライトすべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部と、

複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第二の結合部と、

を備えたことを特徴とするディスク制御システム。

【請求項2】 前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項1に記載のディスク制御システム。

【請求項3】 前記第一の結合部又は前記第二の結合部は、メモリバス用スイッチで構成されることを特徴とする請求項1に記載のディスク制御システム。

【請求項4】 前記第一の結合部は、データ伝送用のケーブルで構成されることを特徴とする請求項1に記載のディスク制御システム。

【請求項5】 前記各ディスク制御装置の内部において、共通の電源から給電される前記各ディスク制御ユニットを前記第一の結合部は結合することを特徴とする請求項1に記載のディスク制御システム。

【請求項6】 複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記第一の結合部を相互に接続することを特徴とする請求項1に記載のディスク制御システム。

【請求項7】 請求項1に記載のディスク制御システムと、前記ディスク制御システムの前記チャネル制御部に対して接続される、データの授受用のホストコンピュータとを備えることを特徴とするディスクシステム。

【請求項8】 請求項1に記載のディスク制御システムと、前記ディスク制御システムの前記ディスク制御部に対して接続される、データ格納用のディスク装置とを備えることを特徴とするディスクシステム。

【請求項9】 複数のディスク制御ユニットを有するディスク制御装置において、

前記ディスク制御ユニットは、

ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、

ディスク装置とのインターフェースを有する一または複数のディスク制御部と、

前記ディスク装置にリード/ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャネル制御部と前記ディスク制御部とを相互に接続する内部結合部と、

を備えており、

前記各ディスク制御装置の内部において、データをリード/ライトすべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部を備えて、前記各ディスク制御ユニットの前記内部結合部は、複数の前記ディスク制御装置に跨り、データを転送すべく、第二の結合部によって、相互に結合されることを特徴とするディスク制御装置。

【請求項10】 前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項9に記載のディスク制御装置。

【請求項11】 前記第一の結合部又は前記第二の結合部は、メモリバス用スイッチで構成されることを特徴とする請求項9に記載のディスク制御装置。

【請求項12】 前記第一の結合部は、データ伝送用のケーブルで構成されることを特徴とする請求項9に記載のディスク制御装置。

【請求項13】 共通の電源から給電される前記各ディスク制御ユニットを前記第一の結合部は結合することを特徴とする請求項9に記載のディスク制御装置。

【請求項14】 複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記第一の結合部が相互に接続されることを特徴とする請求項9に記載のディスク制御装置。

【請求項15】 ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、ディスク装置とのインターフェースを有する一または複数のディスク制御部と、前記ディスク装置にリード/ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャネル制御部と前記ディスク制御部とを相互に接続する内部結合部とを備えた前記ディスク制御ユニットを複数有するディスク制御装置におけるデータ通信の制御方法であって、

前記各ディスク制御装置の内部において、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部により、データをリード/ライトするとともに、

複数の前記ディスク制御装置に跨り、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第二の結

(3)

特開2003-323261

3

4

合部によって、データを転送することを特徴とするディスク制御装置におけるデータ通信の制御方法。

【請求項16】 前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項15に記載のディスク制御装置におけるデータ通信の制御方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、ディスク制御システム、ディスク制御装置、ディスクシステム、及びその制御方法に関する。

【0002】

【従来の技術】半導体記憶装置を記憶媒体とするコンピュータの主記憶のI/O性能に比べて、磁気ディスクを記憶媒体とするディスクサブシステムのI/O性能は3～4桁程度低く、従来からこの差を縮めること、すなわちサブシステムのI/O性能を向上させる努力がなされている。サブシステムのI/O性能を向上させるための1つの方法として、複数の磁気ディスク装置でサブシステムを構成し、データを複数の磁気ディスク装置に格納する、いわゆるディスクシステムと呼ばれるシステムが知られている。このような技術を開示したものとして、特開2001-256003号公報がある。同公報中の図4に示す技術では、スイッチを用いた相互結合網を介して間接的に、ホストコンピュータ50が全てのディスク制御装置4に接続されている。

【0003】しかしながら、複数のディスク制御装置を1つのディスク制御装置として運用するためには、相互結合網を構成するスイッチ内に、そのスイッチに接続された全てのディスク制御装置のデータが、どのディスク制御装置に格納されているかを示すマップを持つ必要があり、ホストコンピュータからアクセス要求があった場合、スイッチにおいてコマンドを解析し、要求データを格納しているディスク制御装置に割り振る機能が必要となる。この場合、従来のチャンネルIF部でのコマンド解析に加え、その上に繋がるスイッチにおいてもコマンドを解析する必要があるため、ホストコンピュータがディスク制御装置に直接接続されている場合に比べ、性能が低下するという問題がある。

【0004】そこで、この特開2001-256003号公報に開示された発明では、同公報の図1や図8に示されるように、相互結合網を介して、全てのチャンネルIF部及びディスクIF部から、全ての共有メモリ部あるいは全てのキャッシュメモリ部へアクセス可能な構成となっている。

【0005】このような技術により、小規模な構成から超大規模な構成まで、同一の高機能・高信頼性のアーキテクチャで対応可能であって、スケーラビリティのある構成のディスク制御装置を提供できる。

【0006】

【発明が解決しようとする課題】しかしながら、前述した従来の技術にあっては、データの転送やリード/ライトの処理の効率が未だ不十分である。場合によっては、アクセスに関し、論理的な競合により、相互結合網の効率が50%以下にまで落ち込んでしまうのである。これを解決しようすると、広帯域化を図る必要があるが、高価格化を招く。

【0007】本発明は、このような課題に鑑みてなされたもので、ディスク制御システム、ディスク制御装置、ディスクシステム、及びその制御方法を提供することを目的とする。

【0008】

【課題を解決するための手段】前記目的を達成すべく、本発明の主たる発明のディスク制御システムでは、複数のディスク制御ユニットを有するディスク制御装置を複数備えたディスク制御システムにおいて、前記ディスク制御ユニットは、ホストコンピュータとのインターフェースを有する一または複数のチャンネル制御部と、ディスク装置とのインターフェースを有する一または複数のディスク制御部と、前記ディスク装置にリード/ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャンネル制御部と前記ディスク制御部とを相互に接続する内部結合部とを備えており、前記各ディスク制御装置の内部において、データをリード/ライトすべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部と、複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第二の結合部とを備えたこととする。

【0009】その他、本願が開示する課題、及びその解決方法は、発明の実施形態の欄及び図面により明らかにされる。

【0010】

【発明の実施の形態】本明細書の記載により、少なくとも次のことが明らかにされる。前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することとしてもよい。

【0011】前記第一の結合部又は前記第二の結合部は、メモリバス用スイッチで構成されることとしてもよい。

【0012】また、前記第一の結合部は、データ伝送用のケーブルで構成されることとしてもよい。

【0013】さらに、前記各ディスク制御装置の内部において、共通の電源から給電される前記各ディスク制御ユニットを前記第一の結合部は結合することとしてもよい。

【0014】さらにまた、複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニ

(4)

特開2003-323261

5

ットの前に第一の結合部を相互に接続することとしてもよい。

【0015】

【実施例】本発明に係る実施例につき、図面を参照して説明する。ディスク制御システム1000は、複数のディスク制御装置200を備えている。各ディスク制御装置200は、複数、好ましくは二つのディスク制御ユニット(DKCとも称する)100(DKC#0乃至DKC#3)を有する。

【0016】これら各ディスク制御ユニット100は、10 チャンネル制御部1と、ディスク制御部2と、DKC内結合部(内部結合部)3とを備える。例えば、DKC内結合部は相互結合網で構成される。チャンネル制御部1は、ホストコンピュータ300とのインターフェースを有する。ディスク制御部2は、ディスク装置400とのインターフェースを有する。DKC内結合部3は、ディスク装置400にリード/ライトされるデータを一時的に格納するキャッシュメモリ(CM)4とチャンネル制御部1とディスク制御部2とを相互に接続する。

【0017】さらに、各ディスク制御ユニット100 20 は、障害対策で電源系統別に二重化されている。ペアの電源クラスタA、Bそれぞれが、チャンネル制御部1と、ディスク制御部2と、DKC内結合部3とを備える。DKC内結合部3は、他方の電源クラスタのチャンネル制御部1及びディスク制御部2、並びに、他方の電源クラスタ側のSW5と接続している。また、ディスク制御部2は、他方の電源クラスタ側のディスク装置400とも接続されている。

【0018】各ディスク制御装置200は、各ディスク制御ユニット100のDKC内結合部3を相互に結合す 30 るSW(スイッチ)5(第一の結合部)を備える。各ディスク制御装置100は、SW5を介し、互いのキャッシュメモリ4にアクセスし、データをリード/ライトするなどの通常のアクセス処理を実行する。

【0019】なお、SW5は、LSI等で構成されるメモリバス用スイッチで構成してもよい。この場合、ディスク制御装置200内の各ディスク制御ユニット100に共通の電源ボックスから給電するとした場合に用いられる。LSI等で構成されるメモリバス用スイッチで構成することで安価にできる。

【0020】あるいは、SW5は、データ伝送用のケーブルで構成してもよい。この場合、ディスク制御ユニット毎に電源ボックスを持たせることで、ディスク制御装置200内の各ディスク制御ユニット100の電源を独立して給電するとした場合に用いられる。各ディスク制御ユニット100の電源を独立して給電することで、電源断に耐え得る構成とできる。

【0021】さらに、ディスク制御システム1000において、複数のディスク制御装置200に跨り、各ディスク制御ユニット100のDKC内結合部3を相互に結 50

6

合するSW(スイッチ)6(第二の結合部)を備える。各ディスク制御ユニット100は、SW6を介し、互いのキャッシュメモリ4にアクセスし、データの転送を実行する。なお、SW6はメモリバス用スイッチで構成してもよい。この場合、ディスク制御システム1000全体におけるディスク制御装置200内の各ディスク制御ユニット100に共通の電源ボックスから給電するとした場合に用いられる。LSI等で構成されるメモリバス用スイッチで構成することで安価にできる。

【0022】なお、変形例として、各ディスク制御ユニット100のDKC内結合部3をSW(スイッチ)6でもって結合するのではなく、各ディスク制御装置200に跨り、各ディスク制御ユニット100のSW5を相互に接続し、通常のアクセスやデータ転送の処理を実行できるようにしてもよい。この場合、SW(スイッチ)6を省略でき、システム構成の簡素化が図れる。

【0023】ここで、前述した、DKC内結合部(内部結合部)3、SW(第一の結合部、密結合)5、及びSW(第二の結合部、疎結合)6の構成に関し、二つの事例を用いてより具体的に説明する。

【0024】====事例1====

本事例1は、SW5及びSW6を同一のプロトコルで実現した事例である。なおかつ、SW5及びSW6はDKC内結合部3を拡張した構成としている。

【0025】まず、LSIで実現したDKC内結合部3のブロック図を図2に示す。図2に示すように、DKC内結合部3は、セレクト部3aとバス制御部3b乃至3iとを備える。このセレクト部3aに対してバス制御部3b乃至3iが接続されている。バス制御部3b、3cは、図1に示される各電源クラスタA、B双方のチャンネル制御部1の接続バスと接続されている。バス制御部3d、3eは、図1に示される各電源クラスタA、B双方のディスク制御部2の接続バスと接続されている。バス制御部3fは、DKC内結合部3の属する電源クラスタAあるいは電源クラスタBのCM4の接続バスと接続されている。バス制御部3g、3hは、図1に示される各電源クラスタA、B側双方のSW5の接続バスと接続されている。バス制御部3iは、図1に示されるDKC内結合部3の属する電源クラスタAあるいは電源クラスタBの側のSW6の接続バスと接続されている。このDKC内結合部3の動作については後述する。

【0026】次に、LSIで構成したSW5のブロック図を図3に示す。なお、SW6のハードウェア構成も図3のSW5と同様である。図1に示すように、SW5は、二つのディスク制御ユニット100における各電源クラスタA、BのDKC内結合部3、即ち、計4つのDKC内結合部3と接続している。したがって、図3のブロック図では、4ポートの入出力を有するSW5の事例が示される。SW5は、4つの制御部5a乃至5dと、受信部5e乃至5hと、送信部5i乃至5lとを備え

(5)

特開2003-323261

7

8

る。各受信部5 e乃至5 hと各制御部5 a乃至5 dとは、リクエスト線及びグラントID線Req/Gntを含んだデータ線で相互に接続されている。また、各制御部5 a乃至5 dは、それぞれ対応する各送信部5 i乃至5 lと接続されている。各受信部5 c乃至5 h及び各送信部5 i乃至5 lは、それぞれバッファを備え、相互に接続されている。

【0027】次に、ケーブルで構成したSW5のブロック図を図4に示す。図4では、図1においてSW5で示された部分をケーブルとして結線した構成を示している。 10

【0028】以上、説明したSW5の動作について、図5のフローチャートを参照して説明する。なお、本明細書のフローチャートにおいて“S”はステップ（工程）を意味する。図5には、SW5を介し、図1に示される二つのディスク制御ユニット100（DKC#0及びDKC#1）のDKC内結合部3間においてデータ及びコマンド等が送受される様子が示される。概念としては、アクセス先のアドレス設定などをディスク制御ユニット100側で行い、直接にアクセス先のアドレスを指定して 20

【0029】具体的には、DKC#0が送信したReadコマンドをSW5はDKC#1へ送信する（S100）。このReadコマンドのデータは、図6（a）に示すように、転送先CMアドレス、転送元CMアドレス、転送長及びコマンドとしてのReadで構成される。次いで、DKC#1のCM4にアクセスがあると、DKC#1はデータとステータスを順に送信する。これらデータとステータスをSW5はDKC#0へ送信する（S110、S120）。このデータには、図6（b）に示すように、転 30

送先CMアドレス、転送元CMアドレス及び転送長が付帯する。

【0030】一方、DKC#0がWriteコマンドとデータを順に送信すると、SW5は、これらWriteコマンドとデータをDKC#1へ送信する（S130、S140）。これらWriteコマンド及びデータのデータ構造は図6（a）（b）に示すものと同様であり、図6（a）におけるコマンドとしてのReadがWriteとなる。これWriteコマンドとデータを受信したDKC#1のCM4にアクセスがあると、DKC#1はステータスを送信する。 40

このステータスをSW5はDKC#0へ送信する（S150）。このステータスのデータ構造は図6（c）に示すものと同様である。

【0031】次に、前述したSW6の動作について、図7のフローチャートを参照して説明する。図7には、SW6を介し、図1に示される二つのディスク制御装置200内のディスク制御ユニット100（DKC#0及びDKC#2）のDKC内結合部3間においてデータ及びコマンド等が転送される様子が示される。概念としては、図1のチャンネル制御部1、ディスク制御部2及びC 50

M4のそれぞれには、その機能を実現するためのプロセッサを備えている。そして、CM4のアドレス管理を行うべく、データ転送に先立ち、DKC#0及びDKC#2のプロセッサ間の通信において、アクセス先のDKCに対してアドレスの設定等を要求し、アクセス先のアドレスを取得する。そして、取得したアクセス先のアドレスを指定してデータ転送を実行する。

【0032】具体的には、DKC#0がデータ転送の要求コマンドを発行し、この要求を受けたSW6はDKC#2へ転送する（S200）。この要求コマンドは、図8（a）に示すように、転送先プロセッサを指定するアドレス、転送元プロセッサを指定するアドレス、転送長、及びコマンドとしての転送要求で構成される。次いで、DKC#2のCM4のアクセスに必要なアドレスを算出し、算出したアドレスを転送許可と共に送信する。これらアドレス及び転送許可を受けたSW6はDKC#0へ転送する（S210）。この転送許可のコマンドは、図8（b）に示すように、転送先プロセッサアドレス、転送元プロセッサアドレス、転送長、及びコマンドとしての転送許可で構成される。次いで、アドレス及び転送許可を受信したDKC#0は、Writeコマンドとデータを順に送信する。すると、SW6は、これらWriteコマンドとデータをDKC#2へ送信する（S220、S230）。これらWriteコマンド及びデータのデータ構造は図6（a）（b）に示すものと同様であり、図6（a）におけるコマンドとしてのReadがWriteとなる。これWriteコマンドとデータを受信したDKC#2のCM4にアクセスがあると、DKC#2はステータスを送信する。このステータスをSW6はDKC#0へ転送する（S240）。このステータスのデータ構造は図6（c）に示すものと同様である。

【0033】====事例2====

本事例2では、SW6については、事例1とは異なり、SW5と異なる別のプロトコルでもって構成する。すなわち、SW6は、例えばホストチャネル同様の接続とし、ファイバチャネル上でマッピングしたSCSIコマンド等により論理アドレスでアクセスすることで実現する。一方、SW5については、事例1同様に、DKC内結合部3を拡張した構成であり、動作も事例1と同様である。したがって、事例1と相違するSW6の構成及び動作を中心に説明する。

【0034】具体的な構成としては、図9のブロック図に示すように、事例1の場合を示す図2のブロック図に比し、バス制御部3 iとSW6接続バスとの間にプロトコル変換部6が挿入されている。この点以外は、前述した図2の場合と同様であるため、図2と相違するプロトコル変換部6について説明する。プロトコル変換部6は、図9に示すように、プロセッサ7 a、メモリ7 b、バス制御部7 c、7 d、バッファ7 e、7 f、及びパケット変換部7 g、7 hで構成される。プロセッサ7 a、

(6)

特開2003-323261

9

10

メモリ7b、バス制御部7c、7d、及びパケット変換部7g、7hは共通のバスに接続される。図9に示すように、メモリ7bを適宜使用するプロセッサ7aの制御の下、バス制御部7c、バッファ7e、パケット変換部7g、バス制御部7dといった順序でプロトコルが変換され、データがDKC内結合部3からSW6へ送信される。反対に、メモリ7bを適宜使用するプロセッサ7aの制御の下、バス制御部7d、バッファ7f、パケット変換部7h、バス制御部7cといった順序でプロトコルが変換され、データがSW6経由でDKC内結合部3へ転送される。

【0035】次に、前述したSW6の動作について、図10のフローチャートを参照して説明する。図10には、SW6を介し、図1に示される二つのディスク制御装置200内のディスク制御ユニット100（DKC#0及びDKC#2）のDKC内結合部3間においてデータ及びコマンド等が転送される様子が示される。DKC#0は、Writeコマンドとデータを順に送信する。すると、SW6は、これらWriteコマンドとデータをDKC#2へ送信する（S300、S310）。このデータアクセスを受けたDKC#2は、SW6経由でステータスを送信する（S320）。

【0036】これらWriteコマンド、データ及びステータスのデータ構造を図11（a）（b）（c）に示す。Writeコマンドは、図11（a）に示すように、転送先ポートアドレス、転送元ポートアドレス、転送長、コマンドとしてのwrite、論理アドレス、及び転送サイズで構成される。データでは、図11（b）に示すように、転送先ポートアドレス、転送元ポートアドレス、及び転送長が付帯する。ステータスは、図11（c）に示すように、転送先ポートアドレス、転送元ポートアドレス、転送長、及びステータス情報で構成される。

【0037】ここで、以上説明した実施例で用いられるSW5の一般的な特性について説明する。図12（a）に示すように、SW5のポート0乃至3の入出力は一对一の関係にあり、データ転送の効率がよい。一方、図12（b）に示すように、SW5のポート0乃至3の入出力が任意の関係の場合では、入力ポート0、1が出力ポート0に対応し、入力ポート2、3が出力ポート2に対応する。この場合、平均的にデータ転送の50%が論理的に競合するため、ハードウェアの性能を50%しか活用できない。このため、データ転送の効率が低下した状態となる。すなわち、図13（b）のグラフに示すように、SW5に接続されるクラスタ（図1中のディスク制御ユニット100に相当）の数が増えるほど、効率が低下することとなる。例えば、図13（b）に示すように複数組のクラスタをSW5に接続すると、効率は50%となる。つまり、効率が低下すると広帯域化が必要となり、高価格化という問題を招く。

【0038】そこで、本発明では、図14のブロック図

に示すように、一つのディスク制御装置200を構成するクラスタ（ディスク制御ユニット100）の数を二つとし、これら2クラスタ間をSW5で接続するとすれば、図13（b）のグラフに示すように、クラスタ数を一組（図中のクラスタ数が“1”の場合）とでき、効率を100%とすることができる。

【0039】次に、以上説明した、ディスク制御システム1000、ホストコンピュータ300、及びディスク装置400を備えたディスクシステムの全体的な動作について、図15及び図16のフローチャートを参照して説明する。適宜、図1のブロック図を参照されたい。なお、図面において“S”はステップ（工程）を意味する。図15に示すように、まず、ホストコンピュータ300が処理の要求を開始する（S10）。ホストコンピュータ300に接続されたディスク制御ユニット100は、自己のキャッシュメモリ4のデータに対するアクセスか否かを判別する（S20）。この判別の結果、自己のキャッシュメモリ4のデータに対するアクセスであれば、そのアクセスパスが正常か否かを確認する（S30）。この確認の結果、アクセスパスが正常であれば、自己のキャッシュメモリ4にアクセスし、データのリード/ライトの処理を実行して終了する（S40→S50）。

【0040】一方、S20において、自己のキャッシュメモリ4のデータに対するアクセスでない場合（S20：NO）、ディスク制御ユニット100は、同じディスク制御装置200内の他方（ペア、電源クラスタA、Bのうちの他方）のディスク制御ユニット100に対するアクセスか否かを判断する（S60）。この判断の結果、他方のディスク制御ユニット100に対するアクセスである場合（S60：YES）には、SW5を介し、他方のディスク制御ユニット100のDKC内結合部3を部を通じて他方のディスク制御ユニット100のキャッシュメモリ4にアクセスする（S70）。

【0041】また、S30において、アクセスパスが正常でない場合にも、S70の処理を実行する。この場合、例えば、電源クラスタA側において、チャネル制御部1からDKC内結合部3を通じたキャッシュメモリ4への通信路に障害が発生した場合、電源クラスタAのチャネル制御部1が電源クラスタBのDKC内結合部3へ接続する。そして、このDKC内結合部3が電源クラスタA側のSW5を介して電源クラスタA内のDKC内結合部3経由で、電源クラスタA内のキャッシュメモリ4にアクセスする。このような迂回ルートを有することで対障害性を向上できる。

【0042】一方、S60における判断の結果、他方のディスク制御ユニット100に対するアクセスでない場合（S60：NO）には、図16のBの処理に移り、他のディスク制御装置200へのデータアクセスと判断し（S80）、SW6を介し、DKC内結合部3経由でキ

(7)

特開2003-323261

11

キャッシュメモリ4へデータアクセスし、データの転送を行う。

【0043】また、本実施の形態にあつては、各ディスク制御装置の内部において、各ディスク制御ユニットの内部結合部を相互に結合する第一の結合部により、データをリード/ライトする。なおかつ、複数のディスク制御装置に跨り、各ディスク制御ユニットの内部結合部を相互に結合する第二の結合部によって、データを転送する。

【0044】第一の結合部による密な結合でもってデータをリード/ライトするとともに、第二の結合部による疎な結合でもってデータを転送する。このような、役割分担された結合方式により、高価格化を招くことなくスケーラビリティを向上できる。

【0045】

【発明の効果】低価格化を維持しながらも、データの転送やリード/ライトの処理の効率並びにスケーラビリティを向上できる。

【図面の簡単な説明】

【図1】 本発明の一実施の形態であるディスクシステムの構成を示すブロック図である。

【図2】 本発明の一実施の形態に係るDKC内結合部3の一構成例を示すブロック図である。

【図3】 本発明の一実施の形態に係るSW5の一構成例を示すブロック図である。

【図4】 本発明の一実施の形態に係るSW5の他の構成例を示すブロック図である。

【図5】 本発明の一実施の形態に係る二つのディスク制御ユニット100間のデータ及びコマンド等がSW5を介して送受される様子を示すフローチャートである。

【図6】 本発明の一実施の形態に係るSW5が受送信するデータ構造の例を示す図表である。

【図7】 本発明の一実施の形態に係る二つのディスク装置200間のデータ及びコマンド等がSW6を介して送受される様子を示すフローチャートである。

【図8】 本発明の一実施の形態に係るSW6が転送す*

12

*るデータ構造の例を示す図表である。

【図9】 本発明の一実施の形態に係るDKC内結合部3及びプロトコル変換部7の一構成例を示すブロック図である。

【図10】 本発明の一実施の形態に係る二つのディスク装置200間のデータ及びコマンド等がSW6を介して送受される様子を示すフローチャートである。

【図11】 本発明の一実施の形態に係るSW6が転送するデータ構造の例を示す図表である。

【図12】 本発明の一実施の形態及び従来の技術に用いられるSW5の一般的な特性を示す模式図である。

【図13】 本発明の一実施の形態及び従来の技術に用いられるSW5を示し、(a)はその接続構成を示すブロック図であり、(b)は接続されるクラスタ数に応じた効率を示すグラフである。

【図14】 本発明の一実施の形態に係るディスク制御ユニット(クラスタ)の接続形態を示すブロック図である。

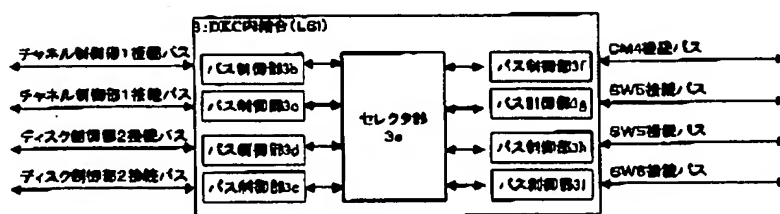
【図15】 本発明の一実施の形態に係るディスクシステムの全体的な動作を示すフローチャートである。

【図16】 本発明の一実施の形態に係るディスクシステムの動作の一部を示すフローチャートである。

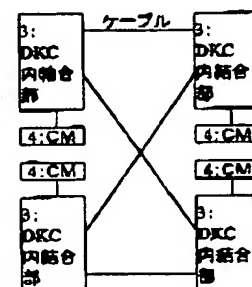
【符号の説明】

- | | |
|------|-----------------|
| 1 | チャンネル制御部 |
| 2 | ディスク制御部 |
| 3 | DKC内結合部 |
| 4 | キャッシュメモリ (CM) |
| 5 | SW (第一の結合部、密結合) |
| 6 | SW (第二の結合部、疎結合) |
| 7 | プロトコル変換部 |
| 100 | ディスク制御ユニット |
| 200 | ディスク制御装置 |
| 300 | ホストコンピュータ |
| 400 | ディスク装置 |
| 1000 | ディスク制御システム |

【図2】



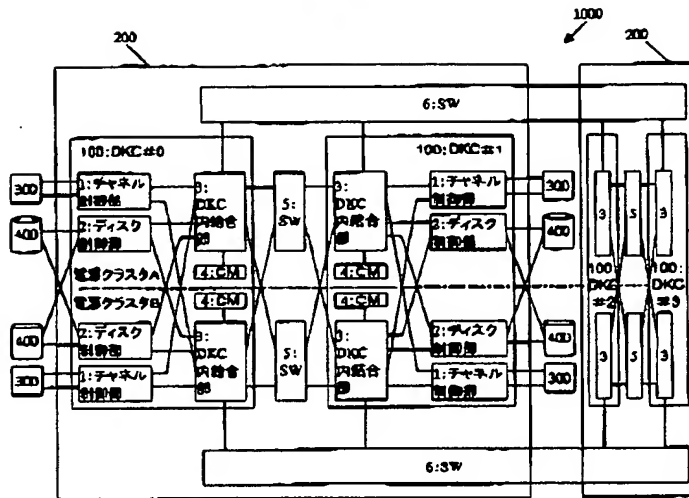
【図4】



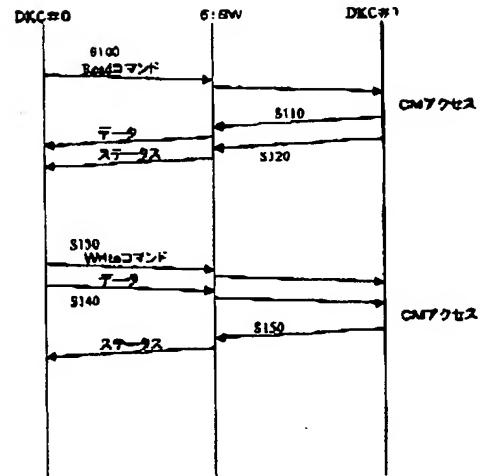
(8)

特開2003-323261

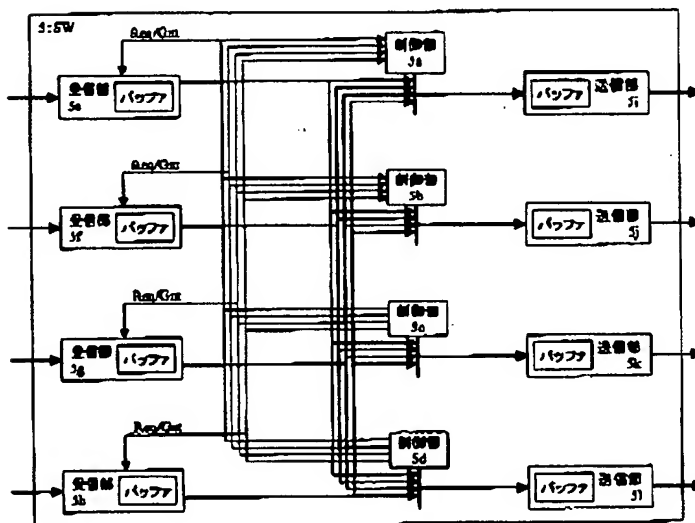
【図1】



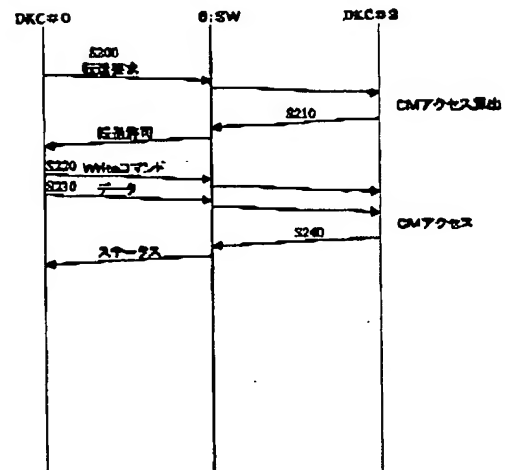
【図5】



【図3】

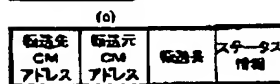
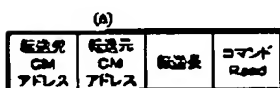


【図7】

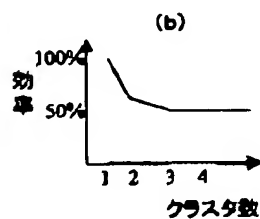
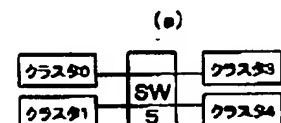
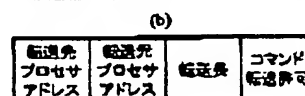
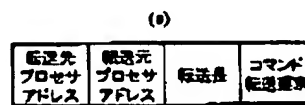


【図13】

【図6】



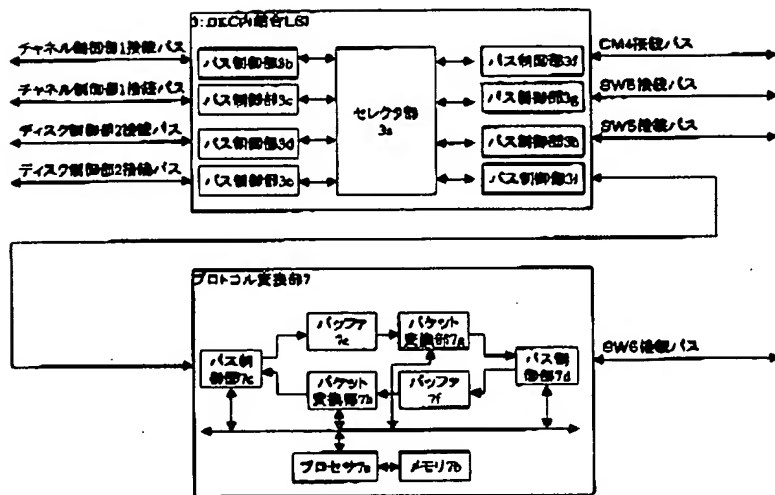
【図8】



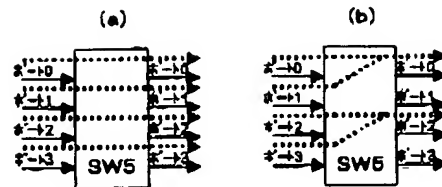
(9)

特開2003-323261

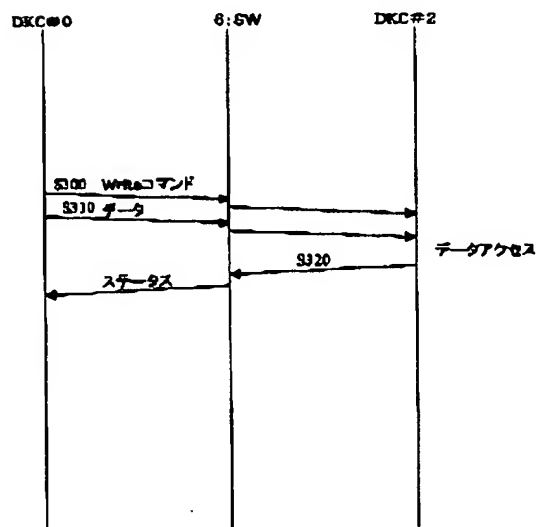
【図9】



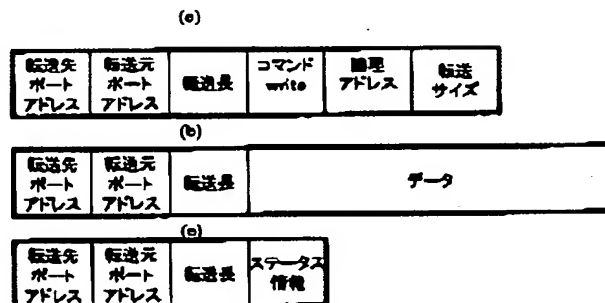
【図12】



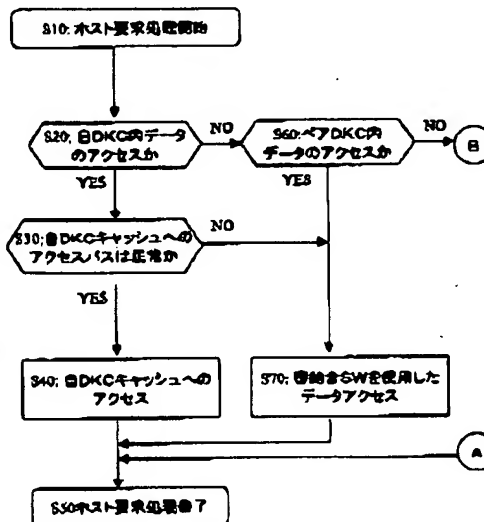
【図10】



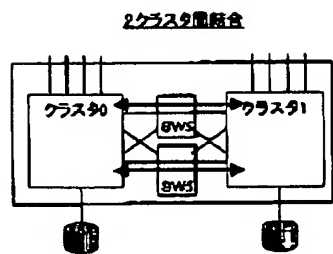
【図11】



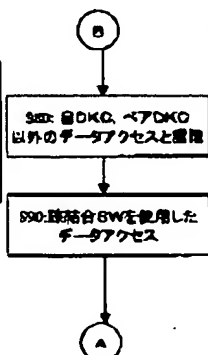
【図15】



【図14】



【図16】



(10)

特開2003-323261

フロントページの続き

(51) Int. Cl.

識別記号

F I

キーワード (参考)

G 0 6 F 13/12

3 1 0

G 0 6 F 13/12

3 1 0 E

Fターム(参考) 5B005 JJ12 MM11 NN75

5B014 EB05 GA13 GA25 GA26 GA47

5B065 BA01 CA07 CE12 CH01 CH11

ZA13

JP 2003-323261 A5 2005.9.22

【公報種別】特許法第17条の2の規定による補正の掲載
【部門区分】第6部門第3区分
【発行日】平成17年9月22日(2005.9.22)

【公開番号】特開2003-323261(P2003-323261A)
【公開日】平成15年11月14日(2003.11.14)
【出願番号】特願2002-126885(P2002-126885)
【国際特許分類第7版】

G 0 6 F 3/06
G 0 6 F 12/08
G 0 6 F 13/12

【F I】

G 0 6 F	3/06	3 0 1 B
G 0 6 F	3/06	3 0 2 A
G 0 6 F	3/06	3 0 2 B
G 0 6 F	12/08	5 0 1 E
G 0 6 F	12/08	5 5 7
G 0 6 F	13/12	3 1 0 E

【手続補正書】

【提出日】平成17年4月7日(2005.4.7)

【手続補正1】

【補正対象書類名】明細書

【補正対象項目名】特許請求の範囲

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

複数のディスク制御ユニットを有するディスク制御装置を複数備えたディスク制御システムにおいて、

前記ディスク制御ユニットは、

ホストコンピュータとのインターフェースを有する一または複数のチャネル制御部と、

ディスク装置とのインターフェースを有する一または複数のディスク制御部と、

前記ディスク装置にリード/ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャネル制御部と前記ディスク制御部とを相互に接続する内部結合部と、

を備えており、

前記各ディスク制御装置の内部において、データをリード/ライトすべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部と、

複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第二の結合部と、

を備えたことを特徴とするディスク制御システム。

【請求項2】

前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項1に記載のディスク制御システム。

【請求項3】

前記第一の結合部又は前記第二の結合部は、メモリバス用スイッチで構成されることを特徴とする請求項1に記載のディスク制御システム。

【請求項4】

前記第一の結合部は、データ伝送用のケーブルで構成されることを特徴とする請求項1

(2)

JP 2003-323261 A5 2005.9.22

に記載のディスク制御システム。

【請求項5】

前記各ディスク制御装置の内部において、共通の電源から給電される前記各ディスク制御ユニットを前記第一の結合部は結合することを特徴とする請求項1に記載のディスク制御システム。

【請求項6】

複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記第一の結合部を相互に接続することを特徴とする請求項1に記載のディスク制御システム。

【請求項7】

請求項1に記載のディスク制御システムと、前記ディスク制御システムの前記チャンネル制御部に対して接続される、データの授受用のホストコンピュータとを備えることを特徴とするディスクシステム。

【請求項8】

請求項1に記載のディスク制御システムと、前記ディスク制御システムの前記ディスク制御部に対して接続される、データ格納用のディスク装置とを備えることを特徴とするディスクシステム。

【請求項9】

複数のディスク制御ユニットを有するディスク制御装置において、
前記ディスク制御ユニットは、
ホストコンピュータとのインターフェースを有する一または複数のチャンネル制御部と、
ディスク装置とのインターフェースを有する一または複数のディスク制御部と、
前記ディスク装置にリード/ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャンネル制御部と前記ディスク制御部とを相互に接続する内部結合部と、
を備えており、
前記各ディスク制御装置の内部において、データをリード/ライトすべく、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部を備えて、
前記各ディスク制御ユニットの前記内部結合部は、複数の前記ディスク制御装置に跨り、データを転送すべく、第二の結合部によって、相互に結合されることを特徴とするディスク制御装置。

【請求項10】

前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項9に記載のディスク制御装置。

【請求項11】

前記第一の結合部又は前記第二の結合部は、メモリバス用スイッチで構成されることを特徴とする請求項9に記載のディスク制御装置。

【請求項12】

前記第一の結合部は、データ伝送用のケーブルで構成されることを特徴とする請求項9に記載のディスク制御装置。

【請求項13】

共通の電源から給電される前記各ディスク制御ユニットを前記第一の結合部は結合することを特徴とする請求項9に記載のディスク制御装置。

【請求項14】

複数の前記ディスク制御装置に跨り、データを転送すべく、前記各ディスク制御ユニットの前記第一の結合部が相互に接続されることを特徴とする請求項9に記載のディスク制御装置。

【請求項15】

ホストコンピュータとのインターフェースを有する一または複数のチャンネル制御部と、ディスク装置とのインターフェースを有する一または複数のディスク制御部と、前記ディ

(3)

JP 2003-323261 A5 2005. 9. 22

スク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリ部と前記チャンネル制御部と前記ディスク制御部とを相互に接続する内部結合部とを備えたディスク制御ユニットを複数有するディスク制御装置におけるデータ通信の制御方法であって、

前記各ディスク制御装置の内部において、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第一の結合部により、データをリード／ライトするとともに、

複数の前記ディスク制御装置に跨り、前記各ディスク制御ユニットの前記内部結合部を相互に結合する第二の結合部によって、データを転送することを特徴とするディスク制御装置におけるデータ通信の制御方法。

【請求項 16】

前記ディスク制御装置は二つの前記ディスク制御ユニットを有しており、前記第一の結合部は、該二つのディスク制御ユニットの前記内部結合部を相互に結合することを特徴とする請求項 15 に記載のディスク制御装置におけるデータ通信の制御方法。

【請求項 17】

複数のディスク制御ユニット及び少なくとも一つの第一の結合部を含む複数のディスク制御装置と、前記ディスク制御装置間に設けられる少なくとも一つの第二の結合部と、を備えるディスク制御システムであって、

前記各ディスク制御ユニットは、ホストコンピュータとのインターフェースを有する少なくとも一つのチャンネル制御部と、ディスク装置とのインターフェースを有する少なくとも一つのディスク制御部と、前記チャンネル制御部と前記ディスク制御部と前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリとを相互に接続する内部結合部と、を含み、

前記各ディスク制御装置の前記第一の結合部は、前記各ディスク制御装置の内部においてデータをリード／ライトすべく、前記各ディスク制御装置を構成する各ディスク制御ユニットの前記内部結合部を相互に接続し、

前記第二の結合部は、前記各ディスク制御装置を跨りデータを転送すべく、前記すべてのディスク制御装置における前記ディスク制御ユニットの前記内部結合部を相互に接続する

ことを特徴とするディスク制御システム。

【請求項 18】

ディスク制御システムと、データを授受するために前記ディスク制御システムと接続される少なくとも一つのホストコンピュータと、を備えるディスクシステムであって、

前記ディスク制御システムは、複数のディスク制御ユニット及び少なくとも一つの第一の結合部を含む複数のディスク制御装置と、前記各ディスク制御装置を跨りデータ転送を行う少なくとも一つの第二の結合部と、を備え、

前記各ディスク制御ユニットは、ホストコンピュータと通信可能なインターフェースを有する少なくとも一つのチャンネル制御部と、ディスク装置と通信可能なインターフェースを有する少なくとも一つのディスク制御部と、前記チャンネル制御部と前記ディスク制御部と前記ディスク装置にリード／ライトされるデータを一時的に格納するキャッシュメモリとを相互に接続する内部結合部と、を含み、

前記各ディスク制御装置の前記第一の結合部は、前記各ディスク制御装置の内部においてデータをリード／ライトすべく、前記各ディスク制御装置のディスク制御ユニットの前記内部結合部の間でデータの転送を行い、

前記第二の結合部は、前記各ディスク制御装置に跨りデータを転送すべく、前記各ディスク制御装置に設けられる全ての前記ディスク制御ユニットの前記内部結合部の間でデータの転送を行い、

前記少なくとも一つのホストコンピュータは、前記いずれかのディスク制御ユニットのチャンネル制御部を介して前記ディスク制御システムに接続される

ことを特徴とするディスクシステム。

【請求項 19】

ディスク制御システムと、データを格納するために前記ディスク制御システムと接続さ

(4)

JP 2003-323261 A5 2005.9.22

れる少なくとも一つのディスク装置と、を備えるディスクシステムであって、

前記ディスク制御システムは、複数のディスク制御ユニット及び少なくとも一つの第一の結合部を含む複数のディスク制御装置と、前記ディスク制御装置間に設けられる少なくとも一つの第二の結合部と、を備え、

前記各ディスク制御ユニットは、ホストコンピュータとのインターフェースを有する少なくとも一つのチャネル制御部と、前記ディスク装置とのインターフェースを有する少なくとも一つのディスク制御部と、前記チャネル制御部と前記ディスク制御部と前記ディスク装置にリード/ライトされるデータを一時的に格納するキャッシュメモリとの間でデータ接続を行う内部結合部と、を含み、

前記第一の結合部は、同一のディスク制御装置内に設けられる前記各ディスク制御ユニットの前記内部結合部を相互に接続して、そのディスク制御装置の内部においてデータをリード/ライトし、

前記第二の結合部は、前記各ディスク制御装置にそれぞれ設けられる前記ディスク制御ユニットの前記内部結合部を相互に接続して、前記ディスク制御装置に跨りデータを転送し、

前記少なくとも一つのディスク装置は、前記いずれかのディスク制御ユニットのディスク制御部を介して前記ディスク制御システムに接続されることを特徴とするディスクシステム。

【請求項20】

複数のディスク制御ユニットと、少なくとも一つの第一の結合部と、を備えるディスク制御装置であって、

前記各ディスク制御ユニットは、ホストコンピュータとのインターフェースを有する少なくとも一つのチャネル制御部と、ディスク装置とのインターフェースを有する少なくとも一つのディスク制御部と、前記チャネル制御部と前記ディスク制御部と前記ディスク装置にリード/ライトされるデータを一時的に格納するキャッシュメモリとを相互に接続する内部結合部と、を含み、

前記第一の結合部は、前記ディスク制御装置の内部においてデータをリード/ライトすべく、前記ディスク制御装置に設けられる前記各ディスク制御ユニットの前記内部結合部間におけるデータ接続を行い、

前記各ディスク制御ユニットの前記内部結合部は、複数のディスク制御装置に跨りデータを転送すべく、少なくとも一つの第二の結合部を介して、他のディスク制御装置の少なくとも一つの内部結合部に接続されることを特徴とするディスク制御装置。

【請求項21】

ホストコンピュータとのインターフェースを有する少なくとも一つのチャネル制御部と、ディスク装置とのインターフェースを有する少なくとも一つのディスク制御部と、前記チャネル制御部と前記ディスク制御部と前記ディスク装置にリード/ライトされるデータを一時的に格納するキャッシュメモリとを相互に接続する内部結合部と、を含む複数のディスク制御ユニットを備えるディスク制御装置におけるデータ通信の制御方法であって、

前記各ディスク制御ユニットの前記内部結合部を相互に接続する第一の結合部を用いて前記ディスク制御装置の内部においてデータをリード/ライトし、

前記各ディスク制御ユニットの前記内部結合部を、他のディスク制御装置の少なくとも一つの内部結合部に接続する第二の結合部を用いて、複数のディスク制御装置を跨りデータを転送する

ことを特徴とするデータ通信の制御方法。